## LESSON ONE: DURABLE, DOABLE DATABASES

## by Frank Sweet

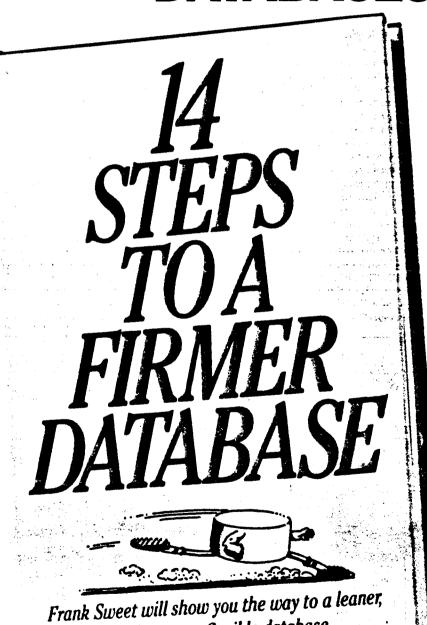
Have you ever noticed that there are just two kinds of data entities in the real world? Imagine, say, a 10,000-record vendor file where, on average, 100 new vendors are added each month and 100 inactive ones purged. Its monthly volatility (turnover divided by population) is 1%. Now consider a file of 1,000 purchase orders where 1,000 orders are added each month and the same number closed. Its volatility would be 100%.

Compute the volatility of the permanent files in your dp shop and you'll unover something strange; there are two populations out there—two classes of records. A histogram shows two peaks: one around 1% like our vendor example, the other near 100% like the purchase orders. Volatility is binary, it seems, and almost every record falls into one group or the other. Why is this? Does it apply to every database? Does it reveal an important underlying truth about data?

The answer to the last two questions is yes. But to find the why, we will journey into the world of database design. We will see that objectively good designs exist, and we'll learn to recognize their traits. We'll review equations in data flow dynamics like the above, and mnemonics like K=DU<sup>3</sup>. Ve'll manipulate the boxes and arrows of Bachman diagrams to derive, split, and inerge entities. We'll examine six useful design patterns we can apply to many applications. And we'll return with a checklist that, applied to a design, tells us if it's ready to be built.

In all, we'll cover 14 topics in the next few months:

- Durable, Doable Databases—the marks of a sound design
- Data Flow Dynamics—the mathematics of data movement
- Objects and Events—the two populations of records
- Keyfield Design—dataless, unique, unambiguous, unchanging



meaner, more flexible database.